A contextual classifier that only requires one prototype pixel for each class

Maletti, Gabriela Mariel; Ersbøll, Bjarne Kjær; Conradsen, Knut

Published in:
Proceedings on IEEE Nuclear Science Symposium Conference Record

Publication date:
2001

Document Version
Publisher's PDF, also known as Version of record

A contextual classifier that only requires one prototype pixel for each class

Gabriela Maletti, Bjarne Ersbøll and Knut Conradsen

Abstract— **A three stage scheme for classification of multi-spectral images is proposed. In each stage, statistics of each class present in the image are estimated. The user is required to provide only one prototype pixel for each class to be seeded into a homogeneous region. The algorithm starts by generating optimum initial training sets, one for each class, maximizing the redundancy in the data sets. These sets are the realizations of the maximal discs centered on the prototype pixels for which it is true that all the elements belong to the same class as the center one. Afterwards a region growing algorithm increases the sample size providing more statistically valid samples of the classes. Final classification of each pixel is done by comparison of the statistical behavior of the neighborhood of each pixel with the statistical behavior of the classes. A critical sample size obtained from a model constructed with experimental data is used in this stage. The algorithm was tested with the Kappa coefficient κ on synthetical images and compared with K-means ($\overline{\kappa} = 0.41$) and a similar scheme that uses spectral means ($\overline{\kappa} = 0.75$) instead of histograms ($\overline{\kappa} = 0.90$). Results are shown on a dermatological image with a malignant melanoma.**

Keywords— **Window Size Optimization, Prototypes, Region Growing, Supervised Classification, Redundancy.**

I. INTRODUCTION

A PREMISE to a supervised classifier is that the training sets provided are statistically valid samples of the classes. Many semi-automatic training and validation set generation schemes by means of region growing algorithms [1], [2], [3] have been proposed. However, these schemes still require some user-input for the estimation of the parameters of the classes. It could be an advantage to develop schemes that minimize the amount of information the user is required to provide. On the other hand, contextual classifiers produce higher classification accuracy. However, the definition of an optimum window size for each step of the image analysis is still a problem [4].

From a semantic and pragmatic point of view [5], a window defines a subset of an entity in which the instances are spatially connected. From an heuristic point of view, the initial optimum window for a given prototype pixel defines the realization of the maximal disc centered on it for which it is true that all the elements belong to the same class as the center one. This can be obtained by detecting the emergence of a state of higher order through the minimization of an optimal learning curve that depends on the fraction of examples [6]. Based on the Heisenberg Uncertainty Principle, it can be deduced that this optimal learning curve corresponds to cases for which the envelope of the signal or the signal itself is a Gauss function. From the Central Limit Theorem we know that the convolution of a large number of positive functions is approximately a Gauss function. Since consecutive values of the function resulting from convolution of two other functions contain a high degree of redundancy [7] we choose as trajectory for the learning curve the points for which the redundancy in the envelope of the fraction of examples is maximized. In the present work, the estimated optimum size of the neighborhood of each prototype pixel is obtained at the minimum of an energy function computed for the most redundant set that can be generated for each window size. This stage will be called "The initial training set generation scheme" [8].

A statistically valid sample of the class can then be obtained by increasing the sample size by means of a "Region Growing" algorithm [9]. Once the samples are defined, the minimum amount of information needed for classifying each pixel in the image has to be established. This is done based on a relation between window size and separability between classes obtained from experimental data, which provides a way of computing the critical number of examples needed for inferring an underlying structure in the data. Pixels are finally classified by comparison of the statistical behavior of its neighborhood with the statistical behavior of the classes [2].

G. Maletti, B. Ersbøll and K. Conradsen are with the Section for Image Analysis and Computer Graphics, Informatics and Mathematical Modelling (IMM), Technical University of Denmark (DTU), 2800-DK Kgs. Lyngby, Denmark. E-mail gmm, be, kc@imm.dtu.dk

II. RATIONALE OF THE ALGORITHM

Let a specific image with B non-correlated bands $X = \{x[r,c]\}$ defined over the given domain $L = \{[r,c]|1 \leq r \leq M, 1 \leq c \leq N\}$ with Q quantization levels. Let K be the number of prototype pixels $x_k^0 \in X$, $1 \leq k \leq K$, seeded into homogeneous regions.

From a syntactic point of view [5], let a window $D_w[r,c]$ define the realization of a disc $\mathsf{D}_w[r,c]$ of radius w centered on the position $[r,c]$ of a given pixel:

$$D_w[r,c] = \{d_w[p] = x[r-i, c-j] | 0 \leq i^2+j^2 \leq w^2 \wedge 1 \leq p \leq n_w\} \quad (1)$$

where n_w is the number of pixels belonging to the set $D_w[r,c]$.

A. The Initial Training Set Generation Scheme

For each prototype pixel x_k^0, an active learning process with increasing window size for each iteration is defined (see left side of Figure 1). It is expected that at some point, the class boundary is detected. This point will be associated to the optimum window size n_{to} for the given pixel. For each window size, the means of the realizations of all the discs of the same size that can be placed into this window are estimated. The estimated variance of the estimated means within groups is then tracked and minimized over the line of maximum redundancy of the mapping model [8].

In general, let the estimation y of the mean of a realization D_w of a disc D_w be the projection [1] of the realization D_w over

[1] Scalar product is denoted by $< ., . >$

0-7803-7324-3/02/$17.00 © 2002 IEEE

the weighting function ω_w:

$$y_w = \sum_p d_w[p]\omega_w[p] \tag{2}$$

where $d_w[p]$ is the p-th pixel belonging to D_w and $\omega_w[p]$ is the weight assigned to it.

Let the realization of a disc $\mathsf{D}_{m,v}[r,c]$ be composed by the estimations $y_v[r-i,c-j]|0 \leq i^2+j^2 \leq m^2$ of the means within n_m groups $D_v[r-i,c-j]$ of size n_v using the same set of weights ω_v. Let $y_{m,v}[r-i,c-j]$ be the estimated mean between these groups using the set of weights ω_m.

Since, for all realization $D_{v,m}\ \exists D_t \wedge \omega_t$ such that $t = t(v,m) \simeq v+m$ and $\omega_t \simeq \omega_v * \omega_m$ (see footnote [2]), the redundancy of a given pair $t(v,m)$ is defined as:

$$R(\mathsf{D}_t) = 1 - \frac{S(\mathsf{D}_t)}{logn_t} \tag{3}$$

where n_t is the number of pixels in the disc D_t and $S(\mathsf{D}_t)$ is the entropy of the disc defined as:

$$S(\mathsf{D}_t) = -\sum_{p=1}^{n_t} \omega_t[p]log(\omega_t[p]) \tag{4}$$

where n_t is the number of pixels of the disc, and $0 < \omega_t[p] < 1$ is the weight assigned to the p-th pixel of the disc.

As can be observed, the measure of redundancy used depends on the group size n_v and the number of groups n_m, and it is inversely related to the entropy of the weights ω_v and ω_m assigned to the elements of the groups and to the groups respectively.

For a given prototype pixel x_k^0 with position $[a,b]$, the size of the initial training set n_{to} corresponds with the minimum estimated variance of the estimated means within groups. This function is computed for each window size for the most redundant set:

$$\mathsf{Z}^2[t(v_o,m_o)] = \sum_{p=1}^{n_{m_o}} \omega_{m_o}||y_{v_o}[p] - y_{v_o,m_o}[a,b]||^2 \tag{5}$$

where (v_o,m_o) is the pair for which the redundancy is maximal for each fixed radius $t(v,m)$, $||\ ||$ is the Euclidian norm, $y_{v_o}[p] \in D_{m_o,v_o}[a,b]$ are the estimated means within groups and $y_{v_o,m_o}[a,b]$ is the estimated mean between groups. The radius t is increased in such a way that the internal to external entropy ratio [8] is smaller than or equal to a constant $R_0\ \forall t$ used. Then

$$t_o = min(\mathsf{Z}^2[t(v_o,m_o)]) \tag{6}$$

Note: In order to simplify the notation, a small change will be done. Starting from here, the notations v and m substitute the notations v_o and m_o of the optimum radius t_o.

B. The Region Growing Algorithm

Once the initial training sets are defined, the sample size can be increased by means of an optimized region-growing algorithm. For each class, pixels satisfying a homogeneity criterion are included in the grown region. This criterion is evaluated in a window of optimum size n_v obtained in the Initial Training Set Generation Stage. A pixel is aggregated into the region if the difference between the homogeneity value of the class and the homogeneity value of the pixel centered window does not exceed a certain threshold. The growing of a region stops, once the homogeneity criterion is no longer satisfied [2]. Following the notation previously introduced, let $G_0 = x^0[a,b]$ be the initial sub-region pointing out a class. In general, $\forall g > 0$ let G'_{g-1} be the set of pixels that do not belong to G_{g-1} but having at least a neighbor with G_{g-1} under a certain connectivity. The set G_g is the region jointly formed by G_{g-1} and the pixels of G'_{g-1} such that the distance from the estimated mean of the class and the estimated mean of the neighborhood centered on those pixels does not differ more than a certain threshold:

$$G_g \equiv \{x[q,r]^g : ||y_v[q,r] - y_{m,v}[a,b]|| \leq \beta Z[t_o]\} \tag{7}$$

where $||\ ||$ is the Euclidian norm and $\beta \in R^+$. The growing of a region continues until $G_{l+1} = G_l$.

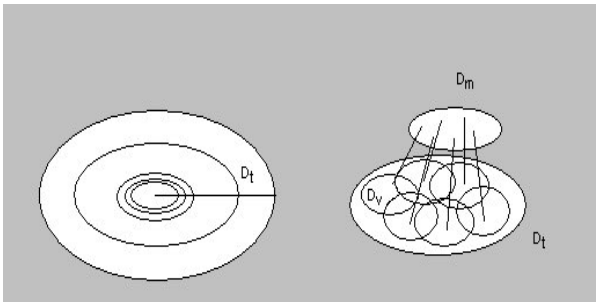

Fig. 1. The mapping model. To the left, a set of discs of increasing radius t centered on a given pixel is shown. To the right, one of these discs is shown in detail: overlapping discs of a fixed radius v are placed into the disc of radius t. A new disc of radius m is constructed with the centers of these overlapping discs of radius v. The mapping model is completely defined when weights are assigned to the elements of each disc.

C. The Final Classification

In this last stage the normalized histograms of the grown regions of each class are used. An optimum window size for image classification is defined using these histograms. Afterwards, the following measure of distance from each pixel to each class is computed; it is the pondered sum of the difference of areas of pairs of class density functions:

$$dA_{(i),(j)} = \frac{1}{2B}\sum_{b=1}^{B}\sum_{n=1}^{Q}|h_{(i)}[b,n] - h_{(j)}[b,n]| \tag{8}$$

where B is the number of bands of the image, Q is the number of quantization levels, $h_{(i)}$ and $h_{(j)}$ are the per band normalized histograms of the i^{th} and j^{th} class respectively.

The critical number of samples for classification is computed as a function of the minimal separability between neighbor classes. This empirical model was constructed doing linear regression of experimental data (See Appendix A).

$$w_c = -6.8341 + \frac{10.1541}{\sqrt{dA}} \tag{9}$$

[2]The equality holds for squares but small differences can be obtained in the case of discs.

where dA has the same meaning as before and w_c is the radius of the estimated optimum disc for classification. It is optimum in the sense that it corresponds to the best classification rate for a set of experimental data with similar statistical descriptions. Figure 2 shows the graphical behavior of this relation.

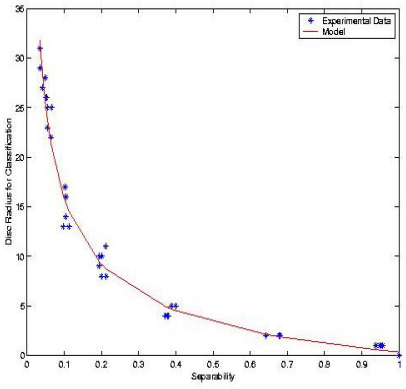

Fig. 2. Separability versus Critical Sample Size

A pixel $x[s,t]$ belonging to the multi-spectral image X is classified according to the following criterion:

$$x[s,t] \rightarrow k : dA_{(x),(k)}, minimum, \forall k \qquad (10)$$

where $_{(k)}$ represents the normalized histogram $h_{(k)}$ of the grown region corresponding to the $k-th$ class and $_{(x)}$ represents the normalized histogram $h_{(x)}$ of the neighborhood of radius w_c centered on the position $[s,t]$ of the pixel x.

III. Results and Discussion

In order to evaluate the algorithm, a set of 15 synthetic images with signal to noise ratios [3] SNR of 0, 3, 8, 15 and 26 and number O (from order) of classes of 3, 5 and 7 was generated. The separation between consecutive means of the classes was 5 gray levels for all images. The classes are Gauss-distributed. The size of the images is 128^2 pixels (See Figure 3). The prototype pixels were seeded in suitable places. The initial optimum window size was estimated for each prototype pixel of each synthetic image setting the Internal to External Entropy Ratio R_o[8] to an empirically found value of 10. The weight assigned to each pixel in each disc was the inverse of the size of the respective disc for all the examples treated. The initial training sets generated are shown in Figure 4. The outputs of the second and third stage of the present scheme are presented in Figures 5 and 6.

The output of the algorithm for this set of synthetical images was tested with Cohen's Kappa coefficient κ. It was compared to K-means and a scheme similar to the present one, except that in the final classification stage, the spectral means of each pixel centered neighborhood are computed and the distance between these means and the estimated means of the classes are used for assigning the pixels to the classes.

Testing the algorithm for dermatological examples introduced some practical considerations that follow. First, the main problem in dermatological images is precisely to define all the classes and their homogeneity. The seed of the prototype pixels is crucial in the sense that this will set the parameters for the

[3]The signal to noise ratio is defined as $SNR = -20[(K-1)\sigma]^{-1} \log \sum_{i=1}^{K-1} |\mu_i - \mu_{i+1}|$ where the μ_i is the mean of the $i-th$ class, the means are ordered in increasing size, and σ is the standard deviation of each one of the K classes

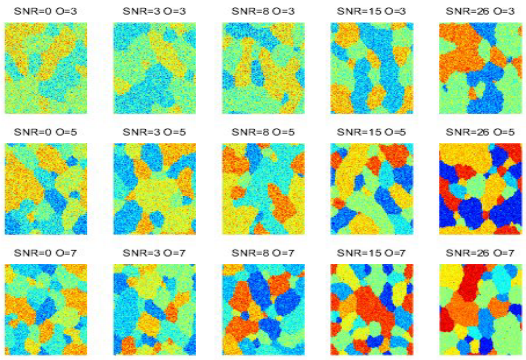

Fig. 3. The original set of synthetical images.

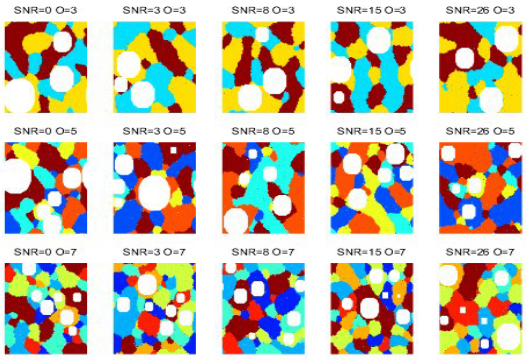

Fig. 4. The initial training sets (in white) centered on the seeded prototype pixels overlayed with the real thematic maps of the synthetical images shown in Figure 3.

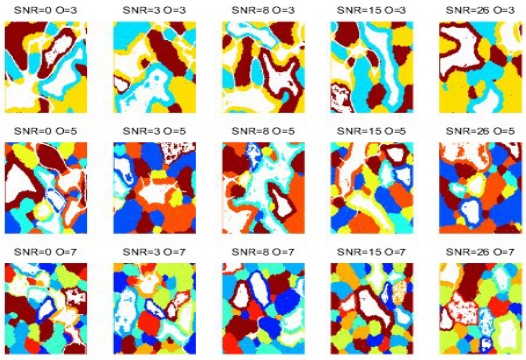

Fig. 5. The grown regions (in white) generated using the initial training sets delineated in Figure 4 overlayed with the real thematic maps of the synthetical images shown in Figure 3.

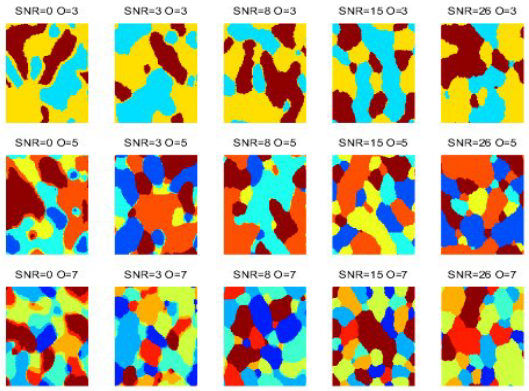

Fig. 6. The estimated thematic maps generated by the present classifier for the set of synthetical images shown in Figure 3 using the training sets delineated in Figure 5.

TABLE I

KAPPA VALUES OF THE DIFFERENT CLASSIFICATION OUTPUTS

SNR	0	3	8	15	26
Order $O = 3$					
K-Means	0.0352	0.0565	0.0911	0.7206	0.9295
Spectral Means	0.6991	0.8718	0.8692	0.8902	0.9565
Norm. Histograms	0.7967	0.9021	0.9164	0.9498	0.9799
Order $O = 5$					
K-Means	0.1186	0.1188	0.3172	0.6964	0.9395
Spectral Means	0.5484	0.5925	0.7812	0.8281	0.8672
Norm. Histograms	0.7350	0.8884	0.9414	0.9583	0.9738
Order $O = 7$					
K-Means	0.1716	0.2537	0.2592	0.4766	0.9347
Spectral Means	0.3995	0.6312	0.6465	0.7225	0.8741
Norm. Histograms	0.7211	0.8511	0.9204	0.9543	0.9749

region growing algorithm. Too homogeneously grown regions will produce classes with almost total separability and, therefore, the final classification will be done with too small window sizes defining not representative real large variation neighborhoods. On the other hand, if the pixels are seeded into regions with large heterogeneity, a high overlap in the statistics of the grown regions can occur, and the final classification stage will have to handle with large neighborhoods in order to make a good discrimination, but border effects can be introduced.

Secondly, since all the pixels are classified, thematic maps produced for real examples can contain miss-classified pixels (healthy skin classified as ill and viceversa). This can be considered as a kind of noise that can be reduced applying an iterative median filter to the map.

An image of a malignant melanoma is the real case presented in this paper. The size of the original image was 885 by 590 pixels. The eigenvalues of the principal components were in descending order: 0.209849, 0.001498 and 0.000137. The two principal components of the original image reduced 50 percent in size were used as input to the algorithm. Five prototype pixels pointing each one a class were seeded into homogeneous regions. The scheme produced a thematic map containing miss-classified pixels (see Figure 8) which number was reduced applying an iterative median filter to the thematic map (see overlay with the original image reduced 50 percent of its size in Figure 9). Pixels with no overlap to one of the classes were assigned to a reject-class.

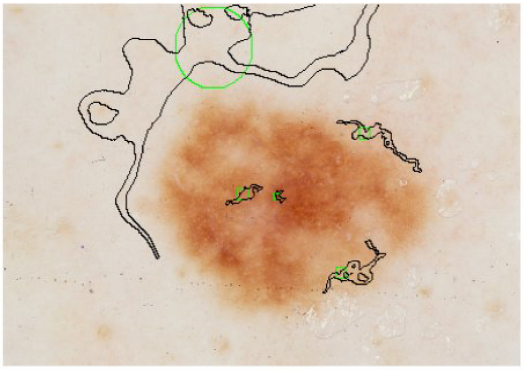

Fig. 7. The initial training sets (delineated in green) centered on the seeded prototype pixels overlayed with the grown regions (delineated in black) of the image of a malignant melanoma.

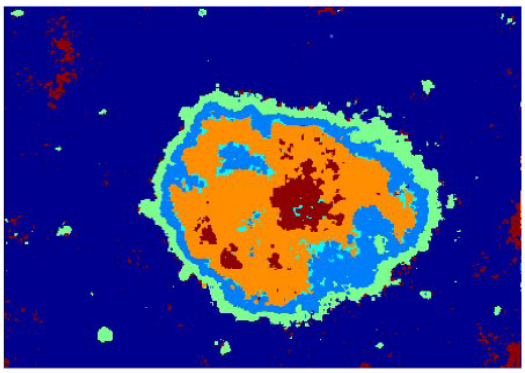

Fig. 8. The thematic map produced by the algorithm for the image.

IV. CONCLUSIONS

A new contextual classifier for multi-spectral images has been developed. The scheme has been tested for synthetical and real examples. It has been shown that the application of the iterative median filter to the thematic map can improve the quality of the results in real examples. In general, results are satisfying both with respect to numerical evaluation and to visual appreciation.

V. ACKNOWLEDGMENTS

To the SITE Project funded by a grant from the Danish Technical Research Foundation (Project Number STVF 56-00-0123) for supporting the present work. To The National Hospital of Denmark for providing the digital images used. To Jorge Lira Chávez (lira@tonatiuh.igeofcu.unam.mx) from the Laboratory of Remote Sensing of the Institute for Geophysics of the National Autonomous University of Mexico for providing source

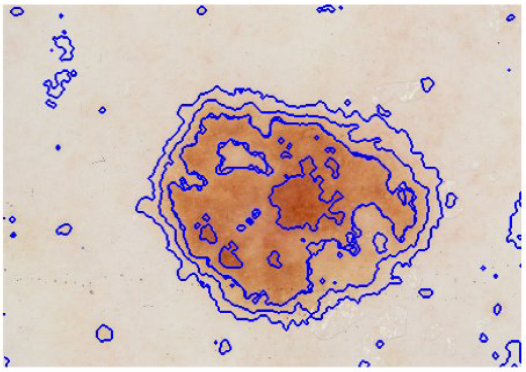

Fig. 9. The original image overlayed with the thematic map after three iterations of the median filter.

code of the package SANDI (Digital Image Analysis System) used in the development of the present scheme.

APPENDIX A

The empirical relation Window size for classification - Separability between Classes.

A set of 36 pairs (A, B) of synthetical images of 128^2 pixels of size was generated. The values of the pixels were generated using different Gauss probability density functions $N(\mu_A, \sigma_A^2)$ and was used $N(\mu_B, \sigma_B^2)$. The values of the parameters were $\sigma_A = \sigma_B = 2^{j-1}, j = 1..6$ and $\mu_A = 128 + \frac{2^{(i-1)}}{2}$ and $\mu_B = \mu_A + \frac{2^{(i-1)}}{2}, i = 1..6$.

For each pair of images (A, B), neighborhoods (discs) of increasing size were centered on the pixels from rows 33 to 96 and columns 33 to 96 of each one of the images. The radius of the discs varied from 0 to 31.

For each window size:

- the normalized histogram of each pixel centered neighborhood was computed;
- the distance dA from this histogram to the normalized histogram of both classes was calculated (See Equation 8) and
- the pixel was assigned to the closest class.
- The quality of the final classification was measured with Cohen's Kappa coefficient κ.

The optimum radius for each pair of images corresponded to the maximum Kappa value one.

The following tables show the distances between each considered pairs of classes and the corresponding optimum disc radii for classification.

TABLE II

THE DISTANCE dA BETWEEN PAIRS OF CLASSES GIVEN BY THE SEPARATION BETWEEN THEIR MEANS $s\mu$ AND THEIR STANDARD DEVIATION σ.

$s\mu/\sigma$	1	2	4	8	16	32
0.5	0.2130	0.1035	0.0511	0.0349	0.0345	0.0549
1	0.3994	0.1998	0.0968	0.0488	0.0411	0.0543
2	0.6416	0.3720	0.1946	0.1022	0.0659	0.0523
4	0.9383	0.6797	0.3879	0.1938	0.1046	0.0644
8	0.9999	0.9493	0.6772	0.3767	0.2111	0.1127
16	1.0000	1.0000	0.9555	0.6797	0.3790	0.2010

TABLE III

THE DISC RADIUS w_c CORRESPONDING TO THE BEST KAPPA VALUE (ONE) FOR EACH PAIR OF CLASSES GIVEN BY THE SEPARATION BETWEEN THEIR MEANS $s\mu$ AND THEIR STANDARD DEVIATION σ.

$s\mu/\sigma$	1	2	4	8	16	32
0.5	8	14	26	29	31	23
1	5	8	13	28	27	25
2	2	4	10	17	25	26
4	1	2	5	9	16	22
8	0	1	2	4	11	13
16	0	0	1	2	4	10

REFERENCES

[1] H. Flesche, A. A. Nielsen, and R. Larsen. Supervised mineral classification with semi-automatic training and validation set generation in scanning electron microscope energy dispersive spectroscopy images of thin sections. *Mathematical Geology*, 32(3):333–366, 2000.

[2] J. Lira and G. Maletti. A supervised classifier for multispectral and textured images based on an automated region growing algorithm. *European Space Agency Publications*, SP-434:153–158, 1998.

[3] J. Lira and L. Frulla. An automated region growing algorithm for segmentation of texture regions in sar images. *International Journal of Remote Sensing*, 19(18):3595–3606, 1998.

[4] M. Hodgson. What size window for image classification? a cognitive perspective. *Photogrammetric Engineering and Remote Sensing*, 64(8):797–807, 1998.

[5] J. Sowa. Peircean fundations for a theory of context. *Lecture Notes in Computer Science*, 1257:41–64, 1997.

[6] A. Buhot and M. Gordon. Phase transitions in optimal unsupervised learning. *Physical Review E*, 57(3):3326–3333, 1998.

[7] M. Møller. Supervised learning on large redundant training sets. *International Journal of Neural Systems*, 4(1):15–25, 1993.

[8] G. Maletti, B. Ersbøll, K. Conradsen, and J. Lira. An initial training set generation scheme. *Proceedings of the* 11^{th} *Scandinavian Conference on Image Analysis.*, pages 706–714, 2001.

[9] J. Lira and G. Maletti. A supervised contextual classification scheme based on an automated region growing algorithm. *5th Iberoamerican Symposium on Pattern Recognition (SIARP 2000), Lisbon, Portugal.*, 11–13 September 2000.